

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

NASH IMPLEMENTATION USING UNDOMINATED STRATEGIES

Thomas R. Palfrey
California Institute of Technology
Center for Advanced Study in the Behavioral Sciences, Stanford /

Sanjay Srivastava
Carnegie-Mellon University



SOCIAL SCIENCE WORKING PAPER 649

December 1986

Nash Implementation Using Undominated Strategies*

by

Thomas R. Palfrey

California Institute of Technology and
Center for Advanced Study in the Behavioral Sciences, Stanford

and

Sanjay Srivastava

Graduate School of Industrial Administration
Carnegie-Mellon University

December 1986

ABSTRACT

This paper provides a characterization of fully implementable outcomes using *undominated Nash equilibrium*, i.e. a Nash equilibrium in which no one uses a weakly dominated strategy. The analysis is conducted in general domains in which agents have complete information. Our main result is that with at least three agents any social choice function or correspondence obeying the usual no veto power condition is implementable unless some players are completely indifferent over all possible outcomes. This result is contrasted with the more restrictive implementation findings with either (unrefined) Nash equilibrium or subgame perfect equilibrium.

* We wish to thank the National Science Foundation for financial support. Professor Palfrey also thanks the Exxon Educational Foundation and the Alfred P. Sloan Foundation for supporting his fellowship at the Center for Advanced Study in the Behavioral Sciences. We have benefitted from comments by seminar participants at Carnegie-Mellon, Penn, Princeton, Stanford, SUNY-Buffalo, and USC, and from discussions with John Moore.

I. Introduction

The implementation problem is to design a game such that a prespecified welfare criterion is guaranteed to be achieved by the game across a large domain of possible environments. As such, the implementation problem is fundamental to economics and its related social science disciplines. The study of this problem combines the essential ingredients of game theory, social choice theory, and the theory of incentives to rigorously analyze the role institutions play in the organization of economic, political and social activity. Because the implementation problem is posed in terms of organizational design, its study has important implications of an applied nature and is directly relevant to important policy problems in regulation, negotiation and bargaining, and contract design. Because it lies at the juncture of several important theoretical subfields, it also provides important insights into the logical foundations of basic theoretical constructs such as equilibrium concepts and social welfare functions.

It is therefore important to solve the implementation problem at a general level, that is, to characterize the boundaries of organizational design: exactly what can be implemented and according to which criteria of rational behavior.

The least controversial notion of rational behavior is given by dominant strategy equilibrium. In such an equilibrium, everyone uses a strategy which is a best response to *any* strategy profile played by others. Clearly, games which possess the property that there exists such an equilibrium across a broad domain of preferences have to be extremely special. This intuition is reflected in the negative results on dominant strategy implementation (Gibbard [1973], Satterthwaite [1975]); Dasgupta, Hammond, and Maskin [1979] provide a summary of results in this area).

These negative results led quite naturally to the widespread use of the much

stronger form of rationality embodied in Nash equilibrium. Nash equilibrium requires that each agent use a strategy which is a best response to the specific strategies being used by the other agents, introducing a form of "strategic rationality" not required by dominant strategy equilibrium. In contrast to the case with dominant strategies, it is quite fairly easy to construct games which possess Nash equilibria across a broad domain of preferences. In fact, with Nash equilibrium, the opposite problem arises : there are frequently too many equilibria. The key to obtaining positive results in Nash implementation theory thus involves mechanism design techniques which eliminate undesirable equilibria.

There have been a number of significant positive results in this area. Hurwicz [1979a], Schmeidler [1980], and Walker [1981] constructed mechanisms which implement Walrasian and Lindahl allocations in Nash equilibria. Maskin [1977] provided a complete characterization of Nash implementable social choice correspondences (SCC's) or welfare criteria. He showed that a condition called monotonicity of an SCC was necessary and essentially sufficient for implementation. It is known that in economic environments, the Pareto correspondence is implementable, as is the constrained Walrasian correspondence. Further, a result of Hurwicz [1979b] shows that under some regularity conditions, any Pareto optimal and individually rational SCC which is Nash implementable must contain the Walrasian SCC. A detailed study of feasibly implementable SCC's in pure exchange environments is given in Hurwicz, Maskin, and Postlewaite [1980], while comprehensive surveys of the area are given by Dasgupta, Hammond, and Maskin [1979], Maskin [1986], Postlewaite [1986], and Groves and Ledyard [1985]. Characterizations of Bayesian-Nash implementable SCC's in economies with incomplete information are given by Postlewaite and Schmeidler [1986], and Palfrey and Srivastava [1985, 1986a].

In this paper, we make two assumptions about rational individual behavior, which combine the features of dominant strategy implementation and Nash implementation. The first assumption, in the spirit of dominant strategy equilibrium, is completely "non-strategic" and simply says that if, regardless of what actions others might be taking, one is never better off and is sometimes worse off taking action A instead of action B, then action A is not taken. In the parlance of game theory, weakly dominated strategies are not used. The second assumption is the familiar best response (Cournot-Nash) criterion. Given what actions others are taking, one chooses an action which does at least as well as any other action.

We then pose the implementation question relative to these assumptions about behavior. What SCC's are consistent with this type of behavior in the sense that there exists an institution under which this type of behavior will exactly reproduce the social choice correspondence? We call this *Undominated Nash Implementation*.

Since the above question has been answered in great detail for the case of Nash implementation, one might wonder why we are bothering to look at *undominated Nash implementation*. One simple reason is that we believe the *undominated criterion* is a reasonable rationality postulate and, as such, should be incorporated in the equilibrium concept. A second reason is that many reasonable welfare criteria fail to be implementable because of an indeterminacy or multiple equilibrium problem associated with Nash equilibrium. That is, the best response criterion provides insufficient restrictions on behavior, the result being that any institution which is to be designed for a robust set of environments will produce too many Nash equilibria in at least some of these environments.

In a recent paper, Moore and Repullo [1986] have pursued an insight also

exploited by Moulin [1979] and Crawford [1979] that refinements of Nash equilibrium permit the implementation of some additional SCCs. They demonstrate that adding a restriction of sequential rationality partially alleviates this multiple equilibrium problem and thereby substantially expands the set of welfare criteria which can be implemented by carefully adding additional "veto" stages to the original game. Their nearly complete characterization of this solution, called *Subgame Perfect Implementation*, has recently been extended by Abreu and Sen [1986].

In addition, Moore and Repullo illustrate several important applications of their result by presenting a series of examples of incentive problems and identifying welfare criteria which cannot be implemented via Nash equilibria, but which can be implemented via subgame perfect equilibria. What they emphasize is that the addition of extra stages accomplishes more than the more familiar method of adding extra strategies in a single stage game. More precisely, the combination of (i) adding extra strategies in an initial stage, (ii) adding extra stages, and (iii) assuming sequential rationality, accomplishes more than only adding extra strategies in a single stage game. What we show in this paper is that if one is willing to concede the condition that weakly dominated strategies are never used in equilibrium, then the addition of extra stages (and the requirement of sequential rationality) is superfluous. Our main result is that with three or more agents, all social choice correspondences satisfying the usual no veto power condition are implementable unless some agents are completely indifferent over all possible alternatives. Furthermore, *many welfare criteria which are implementable via undominated Nash equilibria are not implementable via subgame perfect equilibrium*. Conversely, *if there are more than three players, then any welfare criterion which satisfies no veto power and is implementable via subgame perfect equilibrium is also implementable in a single-stage game using undominated Nash equilibrium*.

The rest of the paper is organized as follows. In the next Section, we set up the model, define and explain the necessary conditions for Nash and subgame perfect implementation. Then we identify a number of important welfare criteria which fail to be implementable in either Nash equilibrium or subgame perfect equilibrium for a variety of different reasons. For each of these examples we provide extremely simple games which successfully implement the desired outcomes using undominated Nash equilibrium. In Section 3, we identify a necessary condition for implementation in undominated Nash equilibrium and also prove that with three or more agents, it is sufficient for implementation of welfare criteria which satisfy the usual no veto power condition identified by Maskin [1977]. In Section 4, we state a number of extensions of this result which involve both the relaxation of NVP and the two person implementation problem. In Section 5, we conclude with our views on the practical and positive implications of a result which appears to place virtually no restrictions at all on the consistency of welfare criteria with noncooperative behavior under conditions of complete information.

II. The Model and Some Examples

II.A The Model

There are I agents, indexed by $i=1,2,\dots,I$. We denote by A the set of alternatives. A state (or an environment), denoted by s , specifies a profile of preferences, one for each agent. S denotes the set of states, and the state is assumed to be common knowledge among the agents. In state s , the preference ordering of agent i is denoted by $R^i(s)$, which is a complete, reflexive, and transitive binary relation on A . $P^i(s)$ denotes the strict preference ordering derived from $R^i(s)$. Since states only distinguish preference profiles, we require that $s \neq s'$ implies $R^i(s) \neq R^i(s')$ for some i . With this convention, each state represents a unique preference profile, and there are no redundant states.

A Social Choice Correspondence (SCC) is a possibly multivalued mapping $F : S \rightarrow A$. For each state of the economy, it specifies a set of alternatives.

A mechanism is a pair (M, g) , where $M = M^1 \times M^2 \times \dots \times M^I$ and g is a function $g : M \rightarrow A$. M^i is the message space of agent i , and g is the allocation rule or outcome function. M^i serves as the strategy set of i at all $s \in S$. Let $M^{-i} = M^1 \times M^2 \times \dots \times M^{i-1} \times M^{i+1} \times \dots \times M^I$, with $m^{-i} \in M^{-i}$.

Definition 1: $m^i \in M^i$ is a best response for i to $m^{-i} \in M^{-i}$ at state s if $g(m^{-i}, m^i) R^i(s) g(m^{-i}, \bar{m}^i)$ for all $\bar{m}^i \in M^i$.

Definition 2: $m = (m^1, \dots, m^I) \in M$ is a Nash equilibrium at s if for all i , m^i is a best response to m^{-i} at state s .

If m is a Nash equilibrium at state s , then $g(m)$ is called an equilibrium outcome at s . Let $NE(s)$ denote the set of all Nash equilibrium outcomes at s .

Definition 3: F is implemented in Nash equilibrium (by the mechanism (M, g)) if for all s , $F(s) = NE(s)$.

We turn next to Nash equilibria which do not involve the use of weakly dominated strategies. We call such equilibria undominated Nash equilibria (see van Damme [1983], p.31).

Definition 4: A Nash equilibrium, m , is weakly dominated at s if there exists i and $m^i \in M^i$ such that

$$g(m^{-i}, m^i) R^i(s) g(m^{-i}, m^i) \text{ for all } m^{-i} \in M^{-i} \text{ and}$$

$$g(m^{-i}, m^i) P^i(s) g(m^{-i}, m^i) \text{ for some } m^{-i} \in M^{-i}.$$

This definition simply states that by playing the alternative strategy m^i , agent i is never worse off relative to playing m^i and he is strictly better off for some strategy combination of the other agents.

Definition 5: $m \in M$ is an undominated Nash equilibrium (UNE) at s if m is a Nash equilibrium at s which is not weakly dominated at s .

Let $UNE(s) = \{ g(m) \mid m \text{ is a UNE at } s \}$ denote the set of undominated Nash equilibrium outcomes at s .

Definition 6: F is implemented in undominated Nash equilibrium (by the mechanism (M, g)) if for all s , $F(s) = UNE(s)$.

II.B Examples

We turn next to a series of examples which illustrate that a number of important SCC's fail to be Nash or subgame perfect implementable.

For Nash implementation, Maskin [1977] identified monotonicity as a very simple and intuitive necessary condition for implementation:

Definition 7: F is monotonic if for all s, s' :

If: (i) $x \in F(s)$

(ii) For all i , $y \in A$, $x R^i(s) y \Rightarrow x R^i(s') y$

Then: $x \in F(s')$

Roughly speaking, this says that if x is in the social choice set at s , and the upper contour set relative to x does not expand for anyone at state s' , then x must also be in the social choice set at s' . To see why this is a necessary condition for Nash implementation is straightforward. Suppose that some game implements F , and $x \in F(s)$ is an equilibrium outcome at s . Since everything preferred to x at s' must also have been preferred to x at s for all agents of the game, the Nash equilibrium which generated x as an outcome at s must still be an equilibrium at s' . Since by assumption the game implements F , this implies that x lies in $F(s')$.

If the domain of F includes a sufficiently large set of possible preference profiles, then monotonicity turns out to be extremely restrictive, the conclusion being that essentially no single valued social choice correspondence (i.e. a social choice function) is Nash implementable. With some fairly strong assumptions on the domain, some more positive results emerge with respect to multivalued correspondences. For example, the ("constrained") Walrasian correspondence is implementable in economic environments. However, in more general environments, even apparently reasonable correspondences are not Nash implementable. The

following example shows this for the Pareto correspondence (This example was pointed out to us by Faruk Gul).

Example 1: Pareto Optimality

There are 3 alternatives, $A=\{x,y,z\}$, 2 states, $S=\{s,s'\}$, and 3 agents. Preferences are given by:

$R^1(s)$	$R^1(s')$	$R^2(s)=R^3(s)=R^2(s')=R^3(s')$
x	xy	y
y	z	x
z		z

The Pareto correspondence evaluated at these two states is $F(s)=\{x,y\}$ and $F(s')=\{y\}$. Unfortunately, monotonicity requires that $x \in F(s')$, and F is therefore not implementable Nash equilibrium. We remark that the Pareto correspondence is monotonic if S is restricted to strict orders on A or if $\{S,A\}$ correspond to neoclassical, pure exchange, economic environments. As a general rule, however, the Pareto correspondence is not monotonic.

Example 2: Plurality Rule (from Abreu and Sen [1986])

There are 3 alternatives, $A=\{x,y,z\}$, 2 states, $S=\{s,s'\}$, and 3 agents. Preferences are given by:

$R^1(s)=R^1(s')$	$R^2(s)=R^2(s')$	$R^3(s)$	$R^3(s')$
x	y	z	y
y	z	y	z
z	x	x	x

Suppose the welfare criterion picks the alternative which is the first choice of the most number of agents, and otherwise uses an arbitrary tiebreaking procedure,

say alphabetical order. Then we get $F(s)=x$ and $F(s')=y$. But monotonicity implies that we must have $x \in F(s')$.

Example 3: Borda Count (and weak majority rule)

There are 3 alternatives, $A=\{x,y,z\}$, 2 states, $S=\{s,s'\}$, and 5 agents. Preferences are given by:

$R^1(s)=R^2(s)$	$R^3(s)$	$R^4(s)=R^5(s)$	$R^1(s')=R^2(s')$	$R^3(s')$	$R^4(s')=R^5(s')$
(2) x	y	z	(2) x	(1) yz	(2) z
(0) y	z	y	(-1) yz	(-2) x	(0) y
(-2) z	x	x			(-2) x

The (adjusted) Borda count (see Black [1958]) for each alternative in each preference ordering is given in parentheses and equals the number of alternatives ranked below the particular alternative in question minus the number of other alternatives ranked above it. This example is set up so that there is a unique Borda winner in each case so $F(s)=y$, $F(s')=z$. Both y and z are also weak majority winners at s and s' respectively. But monotonicity would require that $y \in F(s')$.

Example 4: Majority Rule

There are 5 alternatives, $A=\{v,w,x,y,z\}$, 2 states, $S=\{s,s'\}$, and 3 agents. Preferences are given by:

$R^1(s)$	$R^1(s')$	$R^2(s)=R^2(s')$	$R^3(s)=R^3(s')$
v	v	w	z
y	z	x	x
z	y	y	y
x	x	z	v
w	w	v	w

The form of majority rule we have in mind is that the welfare criterion picks a majority rule winner if one exists (i.e. if there exists an alternative which beats each other alternative in a pairwise vote). If there does not exist a majority rule winner, then the Pareto rule is followed. To avoid the problem illustrated in Example 1, we consider only strict orders. The implied social choice correspondence, evaluated at these two environments is $F(s) = \{v, w, x, y, z\}$ and $F(s') = z$. However, monotonicity implies that $F(s')$ should also contain v, w , and x .

Many other examples can be constructed using more elaborate combinations of majority choice and scoring rules which yield nonmonotonic social choice correspondences in the context of natural (and actually used) performance measures. These include the runoff systems for determining a winner in a multicandidate election when no clear majority winner exists, single and double elimination tournaments, and examples from economic environments such as the ones contained in Moore and Repullo [1986].

In fact the motivation for Moore and Repullo's examples was to show how much more powerful subgame perfect implementation is compared to (unrefined) Nash equilibrium. In the remainder of this section, we demonstrate that none of the correspondences illustrated in examples 1-4 are subgame perfect implementable. We then show via examples in this section and in rigorous generality in the next section that all four of them (and indeed almost any correspondence, including "economic" criteria, Pareto criteria, scoring rules and majority-based schemes) are implementable in undominated Nash equilibrium.

A precise necessary condition for subgame perfect implementation is given by Abreu and Sen [1986]:

Definition 8: F satisfies Condition α if there exists $B \subseteq A$ and $B \supseteq \text{range } F$ such that for all s and s' , if $x \in F(s)$ and $x \notin F(s')$, then

there exists a finite sequence of allocations in B , $a_0 = x$, $a_1, a_2, \dots, a_\ell, a_{\ell+1}$ and a sequence of agents, $j(0), j(1), \dots, j(\ell)$ such that

- (i) $a_k R^{j(k)}(s) a_{k+1}$ for $k=0, 1, 2, \dots, \ell-1$
- (ii) $a_\ell R^{j(\ell)}(s) a_{\ell+1}$ and $a_{\ell+1} P^{j(\ell)}(s') a_\ell$
- (iii) a_k is not $R^{j(k)}(s')$ maximal for $j(k)$ in B , $j=1, 2, \dots, \ell$
- (iv) $j(\ell) \neq j(\ell-1)$ or $\ell=0$, if $a_{\ell+1}$ is $R^i(s')$ maximal in B for all $i \neq j(\ell)$.

Given this condition, it is relatively straightforward to verify whether or not a given F is subgame perfect implementable. Consider Example 1. From part (ii) of condition α it must be the case that since $x \in F(s)$ but $x \notin F(s')$, then there is an agent i and a pair of outcomes (b, c) such that $b R^i(s) c$ and $c P^i(s') b$. Since this is not the case for the SCC used in this example it follows immediately that the Pareto correspondence is not subgame perfect implementable. The following game in which only agents 1 and 2 play, implements this F using undominated Nash equilibrium:

		Agent 2	
		δ	β
Agent 1	δ	y	y
	β	z	x
	γ	y	z

In state s , the pair (δ, δ) and (β, β) are both undominated Nash equilibria. However, in state s' , only (δ, δ) is an undominated Nash equilibrium, since δ is weakly dominated by β for agent 1.

Next, consider Example 2. Abreu and Sen [1986] have shown that this SCC violates Condition α since for all sequences $j(0), \dots, j(\ell)$ and $x = a_0, a_1, a_2, \dots, a_{\ell+1}$, x is $R^{j(0)}(s')$ maximal for $j(0)$. F is therefore not subgame perfect

implementable. We now construct a game which implements F in undominated Nash equilibrium.

		<u>Agent 1</u>	
		δ	β
<u>Agent 3</u>	δ	x	z
	β	x	y
	γ	x	z
		<u>Agent 2 plays δ</u>	

		<u>Agent 2 plays β</u>	
		δ	β
<u>Agent 3</u>	δ	x	z
	β	y	z
	γ	z	y

In the game, agents 1 and 2 each have two strategies, δ and β , while agent 3 has three strategies, δ , β , and γ . Agent 2 chooses a matrix, agent 1 chooses a column, and agent 3 chooses a row. The unique undominated Nash equilibrium at s is (δ, δ, δ) and the unique undominated Nash equilibrium at s' is (β, β, γ) , so that this game implements F .

Next, consider Example 3. For the same reason as in Example 1, F is not subgame perfect implementable. However, the following game implements F in undominated Nash equilibrium :

		<u>Agent 2</u>	
		δ	β
<u>Agent 3</u>	δ	y	y
	β	y	y
		<u>Agent 1 plays δ</u>	

		<u>Agent 1 plays β</u>	
		δ	β
<u>Agent 3</u>	δ	x	z
	β	z	z

Here, agents 4 and 5 do not play the game, agent 1 chooses a matrix, agent 2 chooses a column and agent 3 chooses a row. Here, (δ, β, δ) is the unique UNE at s while (β, δ, β) is the unique UNE at s' , so this game implements F .

Finally, consider Example 4. Here, there does exist an agent and a pair of alternatives for which that agent's preferences are reversed between s and s' as

required by part (ii) of Condition α . However, Condition α requires that $v, w \in F(s')$. To see this for w , consider any pair of sequences $w = a_0, a_1, \dots, a_{\ell+1}$ and $j(0), j(1), \dots, j(\ell)$. We cannot have $j(0) = 2$ since w is $R^2(s')$ maximal. If $j(0) = 1$ or $j(0) = 3$, then part (i) of the condition requires $w R^{j(0)}(s) a_1$ which implies $a_1 = w$. Now we cannot have $j(1) = 2$ for the same reason as above, and the same argument as above for agents 2 and 3 implies $a_2 = a_1 = a_0 = w$. Continuing inductively, there is no pair of sequences satisfying Condition α , so we must have $w \in F(s')$. A similar argument applies to v . We cannot have $j(0) = 1$ since v is $R^1(s)$ maximal. If $j(0) = 2$, then $a_1 = a_0 = v$. If $j(0) = 3$, then $a_1 = v$ or $a_1 = w$. If $a_1 = w$, we have the same problem as in the case of $a_0 = w$. Thus we must have $a_1 = v$, and the argument repeats itself. Thus, F is not subgame implementable, the reason being the lack of a test sequence satisfying Condition α . On the other hand, this example satisfies all the requirements of Theorem 2 of the next Section, and is therefore undominated Nash implementable.

III. Necessary and Sufficient Conditions

This section contains our main results on UNE implementable SCC's. We will show that an extremely weak condition, termed property Q, is necessary for implementation. If there are at least three agents and no veto power holds, then it is also sufficient for implementation.

Definition 7: F satisfies Property Q if for any s, s' , if for all i , $R^i(s) \neq R^i(s')$ implies for all $a, b \in A$, $aR^i(s')b$, then $x \in F(s) \Rightarrow x \in F(s')$.

The condition says that if x is an element of $F(s)$ and the only difference between preferences at s and s' is that at s' some agents are completely indifferent between all alternatives, then x must also lie in $F(s')$.

It is straightforward to see that property Q is necessary for undominated Nash implementation. If x is a UNE outcome at s and preferences at s' are either the same as at s or exhibit complete indifference, then x is also a UNE outcome at s' , so we must have $x \in F(s')$. Formally we state:

Theorem 1: If F is UNE implementable, then F satisfies property Q.

We note that property Q is also a necessary condition for Nash implementation and for subgame perfect implementation.

Property Q is an extremely weak condition, and consequently, it is not surprising that it is necessary for implementation. What is surprising is that with the (usual) additional requirements of three or more agents and no veto power, it is sufficient for UNE implementation. We now turn to this.

The definition of implementation of an SCC has two parts to it. The first part

requires that every element of the SCC be an equilibrium outcome ("truthful implementation"), while the second part requires that every equilibrium outcome be an element of F. To implement an SCC, therefore, we have to ensure that at s , all elements of $F(s)$ are equilibrium outcomes and that nothing outside $F(s)$ ever arises in equilibrium. In the complete information environments we are analyzing, the first part is easy to accomplish; the difficulty arises in ensuring the second part. For example, suppose $x \in F(s)$ but $x \notin F(s')$. Then, we have to ensure that x is an equilibrium outcome at s but not at s' , so that at s' , we have to eliminate x as a potential equilibrium outcome. The role of property Q in eliminating such outcomes is illustrated by the following examples. It is useful for this purpose to re-write property Q as follows.

Definition 9': F satisfies Property Q if for any s, s' , if $x \in F(s)$ and $x \notin F(s')$, then either

(i) there exists i and $a, b \in A$ with $aP^i(s)b$ and $bR^i(s')a$ and there exist $c, d \in A$ with $cP^i(s')d$

or (ii) there exists i and $a, b \in A$ with $aR^i(s)b$ and $bP^i(s')a$.

This states that if $x \in F(s)$ and $x \notin F(s')$ then there must exist some agent with different preferences between the two states and that this agent cannot be completely indifferent between all alternatives at s' . This lack of complete indifference at s' need not be stated explicitly in (i) since the statement $bP^i(s')a$ implies it. Definition 9' is thus equivalent to Definition 9. It is now straightforward to see that (ii) of Condition α implies (ii) of Definition 9' and that Property Q is strictly weaker than Condition α .

We turn next to some examples. All the examples have two states and two agents.

Example 5: $A = \{x, y, z, w\}$, $S = \{s, s'\}$, $I = 2$. Both agents have identical preferences in the two states, given by :

<u>s</u>	<u>s'</u>
x	x
y	y
x	w
w	z

Let $F(s) = \{x, y\}$ $F(s') = \{x\}$. Here, $y \in F(s)$ and $y \notin F(s')$. Part (i) and part (ii) of Definition 9' are both satisfied by agent 1 since $zP^1(s)w$ but $wP^1(s')z$.

The following game implements F:

		<u>Agent 2</u>		
		α	β	γ
<u>Agent 1</u>	α	y	z	w
	β	y	w	x

There are two Nash equilibria to this game at both s and s' : (α, α) yielding y as the outcome and (β, γ) yielding x as the outcome. At s , neither of these equilibria is weakly dominated; for example, playing β does not weakly dominate for player 1 because $zP^1(s)w$. At s' , however, β weakly dominates α for player 1, since now $wP^1(s')z$. The only undominated Nash equilibrium at s' is (β, γ) , yielding x as the unique UNE outcome at s' . F is therefore UNE implementable.

In the above example, we eliminated y as an equilibrium outcome at s' by exploiting a reversal of strict preferences between s and s' . In the next example, we illustrate how a similar type of construction is possible when for some player strict preference at s changes to indifference at s' .

Example 6: $A = \{x, y, z\}$, $S = \{s, s'\}$, $I = 2$. Both agents have identical preferences in each state, given by:

<u>s</u>	<u>s'</u>
x	x
y	y, z
z	

Let $F(s) = \{x, y\}$, $F(s') = \{x\}$. Here, agent 1 satisfies (i) of Definition 9'. The following game implements F in UNE:

		<u>Agent 2</u>	
		α	β
<u>Agent 1</u>	α	y	z
	β	z	x

There are two Nash equilibria to this game at both s and s' , (α, α) and (β, β) . At s , neither (α, α) nor (β, β) is not weakly dominated for either player. At s' , however, β weakly dominates α for player 2 and β weakly dominates α for player 2. Thus, both x and y are UNE outcomes at s while x is the unique UNE outcome at s' , and so F is UNE implementable.

In the next example, we consider the case of an indifference at s becoming a strict preference at s' , which is (ii) of Definition 9'. This turns out to be the most difficult case, and requires the construction of a game with infinite strategy sets for the agents.

Example 7: $A = \{x, y, z, w\}$, $S = \{s, s'\}$, $I = 2$. Both agents have identical preferences in each state, given by :

<u>s</u>	<u>s'</u>
x	x
y	y
z, w	z
	w

Let $F(s) = \{x, y\}$ $F(s') = \{x\}$. Here, $y \in F(s)$ and $y \notin F(s')$. Part (ii) of Definition 9' is satisfied by agent 1 since $wR^1(s)z$ but $zP^1(s')w$. The following game implements F:

		Agent 2							
		α_0	α_1	α_2	α_3	α_4	α_5	α_6
Agent 1	α_0	x	z	z	z	z	z	z
	α_1	z	y	w	w	w	w	w
	α_2	z	y	z	w	w	w	w
	α_3	z	y	z	z	w	w	w
	α_4	z	y	z	z	z	w	w
	α_5	z	y	z	z	z	z	w
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

There are an infinite number of Nash equilibria to this game at both s and s' : (α_0, α_0) yielding x as the outcome and (α_k, α_1) yielding y as the outcome for all $k \geq 1$. At s , none of these equilibria are weakly dominated, since both players are indifferent between z and w . At s' , however, α_{k+1} weakly dominates α_k for player 1, since now z is preferred to w . The only undominated Nash equilibrium at s' is (α_0, α_0) , yielding x as the unique UNE outcome at s' . F is therefore UNE implementable. Note that if this game is truncated at any point, say at α_k for agent 1, then (α_k, α_1) becomes an undominated Nash equilibrium at s' .

These examples illustrate how any change in preferences, except a change to complete indifference, can be used to eliminate undesired equilibria; in all of them, the problem was to eliminate y as an equilibrium outcome at s' . The mechanism constructed below to prove sufficiency is a general formulation of the

intuition behind these examples, and relies heavily on the line of argument used in example 7. The proof also requires the following no veto power condition.

Definition 8: F satisfies no veto power if for all s , whenever $aR^i(s)b$ for all $y \in A$ and for at least $I-1$ agents, $x \in F(s)$.

Theorem 2: If F satisfies property Q and no veto power, and $I \geq 3$, then F is implementable in undominated Nash equilibrium.

Proof: We prove the theorem first for the special case in which in each state, each agent has a best and worst element. At the end, we indicate how the proof changes when this assumption is relaxed. For each i define $b^i: S \rightarrow A$ and $w^i: S \rightarrow A$, such that for all s ,

$$b^i(s) \in B^i(s) = \{x \mid x R^i(s) z \text{ for all } z \in A\} \text{ and}$$

$w^i(s) \in W^i(s) = \{x \mid z R^i(s) x \text{ for all } z \in A\}$. We will call $b^i(s)$ a best element for i at s and $w^i(s)$ a worst element for i at s .

Before proceeding to the sufficiency proof, we need a little more notation. For any s and s' , if part (i) of Definition 9' holds, define $i_1(s, s')$, $y_1(s, s')$, and $y_2(s, s')$ to be the agent and pair of allocations such that $y_1 P^1(s) y_2$ and $y_2 R^1(s') y_1$. If part (ii) of Definition 9' holds, define $i_2(s, s')$, $z_1(s, s')$, and $z_2(s, s')$ to be the agent and pair of allocations such that $z_1 R^1(s) z_2$ and $z_2 P^1(s') z_1$.

Let

$$M^i = M_1^i \times M_2^i \times M_3^i \times M_4^i \times M_5^i \quad \text{and} \quad M = M^1 \times \dots \times M^I$$

$$\text{where } M_1^i = \{ (x, s) \mid x \in F(s), s \in S \}$$

$$M_2^i = S$$

$$M_3^i = \{ -4, -3, 2, -1, 0, 1, 2, \dots \}$$

$$M_4^i = \{ 0, 1, 2, \dots \}$$

$$M_5^i = S$$

To define the allocation rule, g , we divide the message space into a number of regions.

$$D_1 = \{ m \mid m^j = (x, s, s, 0, 0, s) \quad \forall j \}$$

$$D_2 = \{ m \mid \begin{aligned} &\forall j, m^j = (x, s, s, 0, 0, s), \text{ or } m^j = (x, s, s', -1, k^j, s^j) \text{ and } j = i_1(s, s'), \\ &\text{or } m^j = (x, s, s', -3, k^j, s^j) \text{ and } j = i_2(s, s') ; m \notin D_1 \end{aligned} \}$$

$$D_3^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s, 0, 0, s), \text{ or } m^j = (x, s, s', -1, k^j, s^j) \\ &\text{and } j = i_1(s, s'), \text{ or } m^j = (x, s, s', -3, k^j, s^j) \text{ and } j = i_2(s, s') ; \\ &m \notin D_1 \cup D_2 \end{aligned} \}$$

$$D_{4A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -2, 0, s) ; i = i_1(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') \end{aligned} \}$$

$$D_{5A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -2, k^j, s) ; \max_{j \neq i} k^j > 0 ; i = i_1(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') \end{aligned} \}$$

$$D_{6A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s', -2, 0, s) ; \\ &m^i = (x, s, s', -1, k^i, s^i) ; i = i_1(s, s') \end{aligned} \}$$

$$D_{7A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -2, k^j, s) ; i = i_1(s, s'') ; \\ &m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s') ; m \notin D_{6A}^i \end{aligned} \}$$

$$D_{8A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -2, k^j, s) ; i \neq i_1(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s'), \\ &\text{or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') \end{aligned} \}$$

$$D_{4B}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -4, 0, s) ; i = i_2(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s') \end{aligned} \}$$

$$D_{5B}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -4, k^j, s) ; \max_{j \neq i} k^j > 0 ; i = i_2(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s') \end{aligned} \}$$

$$D_{6B}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s', -4, 0, s) ; \\ &m^i = (x, s, s', -3, k^i, s^i) ; i = i_2(s, s') \end{aligned} \}$$

$$D_{7B}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -4, k^j, s) ; i = i_2(s, s'') ; \\ &m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') ; m \notin D_{6B}^i \end{aligned} \}$$

$$D_{8B}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s'', -4, k^j, s) ; i \neq i_2(s, s'') ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s'), \\ &\text{or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') \end{aligned} \}$$

$$D_{9A}^i = \{ m \mid \begin{aligned} &\forall j \neq i, m^j = (x, s, s, i, 0, s) ; \\ &m^i = (x, s, s, 0, 0, s), \text{ or } m^i = (x, s, s', -1, k^i, s^i) \text{ and } i = i_1(s, s'), \\ &\text{or } m^i = (x, s, s', -3, k^i, s^i) \text{ and } i = i_2(s, s') \end{aligned} \}$$

$$D_{9B}^i = \{ m \mid \forall j \neq i, m^j = (x, s, s, i, 0, s) ; m \notin D_{9A}^i \}$$

$$D_{10} = \{ m \mid m \notin D_1 \cup D_2 \cup [\bigcup_i D_3^i \cup \dots \bigcup_i D_{9B}^i] \}$$

For $m \in D_{10}$, let $I^* = \{ i \mid \max \{0, m_3^i\} \geq \max \{0, m_3^j\} \text{ for all } j \}$

and let $i^* = \min_{i \in I^*} (i)$.

For any $i_2(s, s'')$, let $z^*(s^i, s, s'') = \begin{cases} z_1(s, s'') & \text{if } z_1(s, s'') R_1(s^i) z_2(s, s'') \\ z_2(s, s'') & \text{otherwise} \end{cases}$

The allocation rule is defined as follows.

$g(m) =$	x	if $m \in D_1$
	x	if $m \in D_2$
	$w^i(s)$	if $m \in D_3^i$
	$y_1(s, s')$	if $m \in D_{4A}^i$
	$w^i(s'')$	if $m \in D_{5A}^i$
	$y_2(s, s')$	if $m \in D_{6A}^i$
	$b^i(s^i)$	if $m \in D_{7A}^i$ and $k^i > \max_{j \neq i} k^j$
	$w^i(s'')$	if $m \in D_{7A}^i$ and $k^i \leq \max_{j \neq i} k^j$
	$b^i(s^i)$	if $m \in D_{8A}^i$ and $m_3^i = -1$ and $k^i > \max_{j \neq i} k^j$
	$w^i(s'')$	if $m \in D_{8A}^i$ and $m_3^i \neq -1$ or $k^i \leq \max_{j \neq i} k^j$
	$z_1(s, s'')$	if $m \in D_{4B}^i$
	$z_1(s, s'')$	if $m \in D_{5B}^i$
	$z_2(s, s')$	if $m \in D_{6B}^i$
	$z^*(s^i, s, s'')$	if $m \in D_{7B}^i$ and $k^i > \max_{j \neq i} k^j$
	$z_1(s, s'')$	if $m \in D_{7B}^i$ and $k^i \leq \max_{j \neq i} k^j$
	$w^i(s'')$	if $m \in D_{8B}^i$
	$b^i(s)$	if $m \in D_{9A}^i$
	$w^i(s)$	if $m \in D_{9B}^i$
	$b^{i*}(m_5^{i*})$	if $m \in D_{10}$

Discussion of the mechanism:

It is unfortunate that the formal definition of the mechanism appears impenetrable. The mechanism although complicated, is actually quite intuitive. Therefore, as a prelude to the formal proof, we attempt to make this intuition accessible by informally describing the basic structure of the message space and the allocation rule.

A. Message space, M :

The message space has five components. The first component, (x, s) , consists of a state and a permissible (by F) social choice at that state. If F were single valued, and we think of i 's "type" as a conditional distribution on S , then a "direct" mechanism would use a message space identical to this first component. The mechanism will be rigged so that in equilibrium agents must essentially always agree to the state. The second and third components are used jointly by agents to "object" to the state that others are claiming, or alternatively, to approve of someone's objections. The messages $\{-1, -3\}$ in M_3^i are "objection flags" and $\{-2, -4\}$ are "objection approval flags." If i raises a -1 flag, then $m_1^i = (x, s)$ and $m_2^i = s'$ must be such that $i = i_1(s, s')$. If i raises a -3 flag, then m_1^i and m_2^i must be such that $i = i_2(s, s')$. Furthermore, when agent j sends a -2 flag and reports $m_1^j = (x, s)$ $m_2^j = s'$, he only gives permission to $i_1(s, s')$ to raise a -1 flag. Similarly j 's partial message $(x, s, s', -4)$ only gives $i_2(s, s')$ permission to raise a -3 flag. The positive integers in M_3^i are used to create a "rat-race" when there is disagreement among some players over M_1 , or if some players are sending otherwise inappropriate messages. The latter can occur for example if i 's message is $(x, s, s', -1, \dots)$ but $i \neq i_1(s, s')$. This rat-race has no equilibrium, since each player wishes to report a higher integer than the other players.

The integers in D_4 are used to create a "tail-chasing" phenomenon for an agent. That is, an agent always wishes to say a higher integer than himself. This feature was illustrated earlier in example 7. The last component of the message space, M_5^i , is used out of equilibrium in conjunction with M_4^i to make sure that i wants to chase his own tail. If he happens to say the highest M_4 integer, then he will receive a best element evaluated at M_5^i (i.e. $b^i(m_5^i)$).

In summary, the regions of the message space are defined as follows. Region 1 is the "equilibrium" region; here, there is total agreement and typical UNE's lie in this region. In region 2, at least one player has raised an objection flag but no permission flags have been raised; otherwise, there is total agreement. Regions 3 to 9B are indexed by i . Region 3 consists of illegitimate unilateral deviations by i from total agreement. In regions 4A to 8A all $j \neq i$ have given i permission to make a -1 objection. Similarly, in regions 4B to 8B all $j \neq i$ have given i permission to make a -3 objection. In region 9, agents $j \neq i$ jointly grant i the right to choose between his most preferred outcome (9A) and his least preferred outcome (9B). Finally, in region 10, the agent with the highest M_4 integer chooses his most preferred outcome.

B. Allocation rule, g :

The mechanism is set up so that it has the following intuitive features:

- (1) Equilibria always involve total agreement (generically $(x, s, s, 0, 0, s)$).
- (2) Objections by i are ignored unless everyone else gives i approval.
- (3) Unilateral deviations from total agreement (other than appropriate objections) are punished.
- (4) Approved objections lead to either "rat-races" or "tail-chasing."
- (5) Unapproved objections (D_2) lead to "tail-chasing."

In order to prove the theorem, we have to show that for all s , $F(s) = \text{UNE}(s)$. This is done in the following Lemmas. The first two Lemmas show that for any s and for any $x \in F(s)$, $m^j = (x, s, s, 0, 0, s) \forall j$ is a UNE at s , leading to the conclusion that for all s , $F(s) \subset \text{UNE}(s)$. The first Lemma shows that such a strategy is a Nash equilibrium at s .

Lemma 1: At s , for any $x \in F(s)$, $m^j = (x, s, s, 0, 0, s)$ for all j is a Nash equilibrium.

Proof: Any unilateral deviation by i moves the aggregate message from D_1 to either D_2 or D_3^i . In either case, i is no better off. ■

The next lemma shows that the strategy described in Lemma 1 is not weakly dominated at s for any agent. This is somewhat complicated by the fact that we need to examine what happens for every possible strategy of the other agents. Regions D_{4A}^i and D_{4B}^i are used to ensure that at s , there does not exist s' such $i=i_1(s, s')$ will deviate by saying $(x, s, s', -1, k^i s^i)$; regions D_{4B}^i to D_{6B}^i are used to ensure that at s , there does not exist s' such that $i=i_2(s, s')$ will deviate by saying $(x, s, s', -3, k^i s^i)$; regions D_3^i , D_{9A}^i and D_{9B}^i are used to ensure that no other type of agent will deviate in any way.

Lemma 2: At s , for any $x \in F(s)$, $m^j = (x, s, s, 0, 0, s)$ for all j is not weakly dominated.

Proof: See Appendix.

Lemmas 1 and 2 yield that for all s , $F(s) \subset \text{UNE}(s)$. We now show that for all s , $\text{UNE}(s) \subset F(s)$. The next Lemma examines the possibility of equilibria outside D_1 and D_2 , and uses no veto power to ensure that any such equilibrium will produce outcomes in the social choice correspondence. The region D_{10} plays an essential role in this argument. Note that in D_{10} , any agent reporting m_3^i which is strictly

positive and strictly greater than $m_3^j \forall j \neq i$ can obtain his most preferred alternative at the true state.

Lemma 3: At any s , if m is a UNE and $m \notin D_1 \cup D_2$, then $g(m) \in F(s)$.

Proof: If m is a Nash equilibrium and $m \notin D_1 \cup D_2$, then we claim that the hypothesis of no veto power is satisfied. To see this, observe that outside of $D_1 \cup D_2$, there are always at least $I-1$ agents each of whom can unilaterally move the aggregate message to D_{10} and obtain their best element at s . Hence, if m is a Nash equilibrium and $m \notin D_1 \cup D_2$, it must be the case that $g(m)R^i(s)b^i(s)$ for at least $I-1$ agents. No veto power then yields $g(m) \in F(s)$. Since $UNE(s) \subset NE(s)$, the conclusion follows. ■

Lemma 3 takes care of all equilibria outside D_1 and D_2 , and we turn next to possible equilibria in D_1 and D_2 . The next Lemma considers the set of equilibria in D_1 . It shows that if $x \in F(s)$ is an equilibrium outcome at s' , then $x \in F(s')$. We show that if $x \notin F(s')$, then $i_1(s, s')$ or $i_2(s, s')$ should deviate by saying $(x, s, s', -1, k^i, s')$ or $(x, s, s', -3, k^i, s')$ respectively.

Lemma 4: If $m = (x, s, s, 0, 0, s) \forall j$ is a UNE at s' , then $x \in F(s')$.

Proof: See Appendix.

Finally, we examine possible equilibria in D_2 . The argument here relies heavily on the idea behind Example 7; we essentially argue that any strategy of the type $(x, s, s', -1, k^i, s')$ is weakly dominated by $(x, s, s', -1, k^i+1, s')$ and that any strategy of the type $(x, s, s', -3, k^i, s')$ is weakly dominated by $(x, s, s', -1, k^i+1, s')$.

Lemma 5: Suppose $m \in D_2$ is a UNE at s' with $m_1^j = (x, s) \forall j$. Then, $x \in F(s')$.

Proof: See Appendix.

To complete the proof of the Theorem, Lemmas 1 and 2 show that $F(s) \subset UNE(s)$, while Lemmas 3-5 show that if m is a UNE at s , $g(m) \in F(s)$, i.e. $UNE(s) \subset F(s)$. These yield $F(s) = UNE(s) \forall s$.

Finally, we indicate how the mechanism has to be modified when some or all agents have no best or worst elements at some states.

If agent i has no worst element, the outcomes in the various regions have to be changed as follows: in D_3^i and D_{9B}^i to something strictly worse than x at s preferences; in D_{5A}^i and D_{7A}^i to something strictly worse than both $y_1(s, s')$ and $y_2(s, s')$ at s' preferences; in D_{8A}^i , $w^i(s')$ should be replaced by x if $i \neq i_2(s, s')$ and $z_1(s, s')$ if $i = i_2(s, s')$.

If agent i has no best element, then $y^i(s^i)$ should be replaced by something strictly better than $w^i(s^i)$ at s^i preferences in D_{7A}^i and D_{8A}^i . In D_{9A}^i , $y^i(s)$ should be replaced by something strictly better than the outcome in D_{9B}^i at s preferences. In D_{10} , define a sequence $\{b_n\}$ with $y_{n+1} P^{i*}(m_5^{i*}) y_n \forall n$, and let the outcome be y_n where $n = m_4^{i*}$.

This concludes the proof of the theorem. ■

IV. Discussion and Extensions

In this section, we consider extensions of Theorem 2. Our first extension deals with relaxing the no veto power condition. While no veto power is a very weak restriction (for example, it is vacuously satisfied in pure exchange economies), it can sometimes impose undesirable restrictions on a welfare criterion. This is illustrated by the following example, which shows that no veto power can be inconsistent with Pareto optimality, and also that it is not necessary for UNE implementation.

Example 8: (No veto power and Pareto optimality)

There are 3 alternatives, $A = \{x, y, z\}$, 2 states, $S = \{s, s'\}$, and 3 agents. Preferences are given by:

$R^1(s)$	$R^1(s')$	$R^2(s) = R^2(s') = R^3(s) = R^3(s')$
x	x, y	x, y
y	z	z
z		

The Pareto optimal SCC for this example is given by $F(s) = \{x\}$, $F(s') = \{x, y\}$. No veto power requires that $y \in F(s)$, so that F does not satisfy the requirements of Theorem 2. However, F is implemented in undominated Nash equilibrium by the following trivial game:

<u>Agent 1</u>	α	x
	β	y

At s , α is the unique undominated Nash equilibrium yielding x as the equilibrium outcome while at s' , both α and β are undominated Nash equilibria yielding x and y as UNE outcomes.

The following assumptions, which are weaker than no veto power, apply to situations similar to the one above.

Weak NVP: For all s, s' , for all i , $W^i(s) \cap [\bigcap_{j \neq i} B^j(s')] \subset F(s')$.

Unanimity: $\bigcap_j B^j(s') \subset F(s')$

Weak NVP says that if a worst element for i at s is a best element at s' for all $j \neq i$, then this element lies in $F(s')$. Unanimity says that any unanimously best element at s' must also lie in F . Together, these assumption are clearly weaker than no veto power.

Theorem 3: If F satisfies Property Q, Unanimity and Weak NVP, and $I \geq 3$, then F is implementable in undominated Nash equilibrium.

Proof: Same as that of Theorem 2 with the exception of Lemma 3. In Lemma 3, weak NVP yields that if there is an equilibrium at s in D_3^i for some i , then the outcome lies in $F(s)$. Part (ii) yields that if there is an equilibrium at s in $D_{4A}^i - D_{10}^i$, then the outcome lies in $F(s)$. ■

We turn next to the case of two agents. In this case, we show that the additional (domain) assumption that there exists a socially undesirable "holocaust" alternative suffices for implementation. Such an assumption has been used by McKelvey [1985] to obtain strategy space reductions for Nash implementation, and by Moore and Repullo [1986] for subgame perfect implementation in the 2 agent case. We also show that with this assumption, no veto power can be replaced with unanimity.

Assumption H: There exists $w \in A$ such that $aP^i(s)w$ for all $a \in A$, for all i , for all s , and $w \notin F(s)$ for all s .

Theorem 4: If F satisfies Property Q, Unanimity, H, and $I \geq 2$, then F is implementable in undominated Nash equilibria.

Proof: If $I > 2$, then Theorem 3 applies. If $I = 2$, we modify the mechanism used in the proof of Theorem 2. Note first that with $I=2$, D_3^i and D_3^j may have a non-empty intersection. We let $g(m) = w$ if $m \in D_3^i$ for any i . Second, since $x \in F(s)$ is never a worst element for any i and s , we let the outcome in D_{9A}^i and D_{9B}^i be determined as in D_{10} . It can now be checked that Lemmas 1,2,3, and 5 still hold. In Lemma 3, $m \in D_3^i$ is never an equilibrium outcome at any s since either agent can unilaterally move m to $D_1 \cup D_2$ and get $x \succ^i(s)w$. Finally, unanimity ensures that if m is an equilibrium at s in $D_{4A}^i - D_{10}$, $g(m) \in F(s)$. ■

V. Discussion and Conclusions

Our results show that in general collective choice problems, virtually all welfare criteria are implementable in undominated Nash equilibrium. In this section we comment on some features of our analysis.

Undominated Nash equilibrium is a refinement of the Nash equilibrium concept which is weaker than Selten's [1975] notion of a (trembling hand) perfect equilibrium (see van Damme [1983] for a discussion of this point). One advantage of UNE over perfection is that it is easy to define for infinite games, while extensions of perfect equilibrium to infinite games can lead to weakly dominated strategies being played in equilibrium. In our setting, it is natural to work with infinite games. If the set of alternatives, A , is infinite (as would be the case, for example, if A was the set of feasible allocations in a pure exchange economy), then the set of preferences over A is also infinite, and any game which involves elicitation of preferences will naturally involve an infinite strategy space. In this sense, UNE is a more natural concept of equilibrium for our setting than is perfection. On the other hand, there are games for which UNE outcomes can appear "unreasonable" relative to perfect equilibrium outcomes. This does not pose a problem for our results since our game is constructed to have only "reasonable" UNE outcomes, where "reasonable" is specified by the welfare criterion. This argument also applies to any further refinement of Nash equilibrium. It should also be noted that in the characterization of perfect equilibrium in term of lexicographic Nash equilibrium (see Brandenburger and Dekel [1986], Blume [1986]), perfection requires strong assumptions on "higher order" beliefs while the relaxation of these assumptions yields undominated Nash equilibrium.

Our next comment concerns the size of the mechanism we construct. This mechanism may appear to be unintuitive and complicated. These complications are

due to the fact that we constructed a single mechanism to implement any SCC satisfying our conditions. Consequently, we were not able to use information which might be specific to a given problem. For example, the mechanism constructed in Palfrey and Srivastava [1986b] for pure exchange economies is much simpler than the one constructed here, but uses several important features of the pure exchange model. Further, as the examples in Section II show, in many cases the implementing mechanism is quite simple.

We turn next to the implications of our results. The results provide a general possibility theorem and show that noncooperative behavior is consistent with an extremely large class of welfare criteria. Since part of the motivation for our analysis was that previously studied equilibrium notions were unable to implement several important welfare criteria, this is a strong positive result insofar as UNE allows us to implement most SCC's. On the other hand, our results also show that noncooperative behavior imposes virtually no restrictions on outcomes attainable as equilibria to games, implying that the theory has very little predictive power. In this light, our results may be construed as being negative in nature.

Finally, we note that we have limited the analysis to situations of complete information. While no general results are as yet available for the case of incomplete information, a characterization of undominated Bayesian implementable allocations in pure exchange economies is given in Palfrey and Srivastava [1986c].

APPENDIX

Proof of Lemma 2:

Consider any i . There are three types of deviations i can make:

- (i) $\bar{m}^i = (x, s, s', -1, k^i, s^i)$ where $i = i_1(s, s')$ for some s'
- (ii) $\bar{m}^i = (x, s, s', -3, k^i, s^i)$ where $i = i_2(s, s')$ for some s'
- (iii) any other deviation.

Case (i): Consider $\bar{m}^j = (x, s, s', -2, 0, s) \forall j \neq i$. Then, $(\bar{m}^{-i}, \bar{m}^i) \in D_{4A}^i$ while $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i$, so $g(\bar{m}^{-i}, \bar{m}^i) = y_1(s, s')$ and $g(\bar{m}^{-i}, \bar{m}^i) = y_2(s, s')$. Therefore if i plays \bar{m}^i instead of m^i , the outcome changes from $y_1(s, s')$ to $y_2(s, s')$ when all other agents play \bar{m}^j . By definition of y_1 and y_2 , $y_1 P^i(s) y_2$, so that agent i is strictly worse off in this case. Hence, \bar{m}^i does not weakly dominate m^i .

Case (ii): First note that if $(x, s, s', -3, k^i, s^i)$ weakly dominates m^i , then so does $(x, s, s', -3, k^i, s)$, because the last element of i 's message is only used in D_{7B}^i to determine z^* and in D_{10} to determine i 's best element when $i = i^*$. In either case, i is at least as well off playing $(x, s, s', -3, k^i, s)$ as by playing $(x, s, s', -3, k^i, s^i)$. Without loss of generality, therefore, we let $\bar{m}^i = (x, s, s', -3, k^i, s)$. For any \bar{m}^{-i} , consider $g(\bar{m}^{-i}, m^i)$ and $g(\bar{m}^{-i}, \bar{m}^i)$, and observe that the outcome is only affected when $(\bar{m}^{-i}, m^i) \in D_{4B}^i \cup D_{5B}^i$. If $\bar{m}^j = (x, s, s'', -4, 0, s) \forall j \neq i$ and $i = i_2(s, s'')$ and $s'' \neq s'$, the outcome can change from $z_1(s, s'')$ in D_{4B}^i to either $z^*(s, s, s'')$ or $z_1(s, s'')$ in D_{7B}^i . By definition of $z_1(s, s'')$ and $z_2(s, s'')$, $z_1 R^i(s) z_2$, so $z^*(s, s, s'') = z_1(s, s'')$, and the outcome therefore does not change. If $\bar{m}^j = (x, s, s', -4, 0, s) \forall j \neq i$, $(\bar{m}^{-i}, m^i) \in D_{4B}^i$ and $(\bar{m}^{-i}, \bar{m}^i) \in D_{6B}^i$ so the outcome changes from $z_1(s, s')$ to $z_2(s, s')$. However, $z_1(s, s') R^i(s) z_2(s, s')$ so i is no better off. If $(\bar{m}^{-i}, m^i) \in D_{5B}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$ and the outcome is

either $z^*(s, s, s'')$ or $z_1(s, s'')$ in D_{7B}^i . Again, since $z^*(s, s, s'') = z_1(s, s'')$, the outcome does not change. We conclude that $g(\bar{m}^{-i}, m^i)R^i(s)g(\bar{m}^{-i}, \bar{m}^i)$ for all \bar{m}^{-i} , so that \bar{m}^i does not weakly dominate m^i .

Case iii: Any other deviation by i moves the aggregate message from D_1 to D_3 .

There are two cases : $xP^i(s)w^i(s)$ and $w^i(s)R^i(s)x$. If $xP^i(s)w^i(s)$ then i is strictly worse off by switching from m^i . If $w^i(s)R^i(s)x$, so that x is a worst element for i at s , consider $\bar{m}^j = (x, s, s, i, 0, s) \forall j \neq i$. Then, $(\bar{m}^{-i}, m^i) \in D_{9A}^i$ while $(\bar{m}^{-i}, \bar{m}^i) \in D_{9B}^i$, and the outcome changes from $b^i(s)$ to $w^i(s)$. If $b^i(s)P^i(s)w^i(s)$ then i is strictly worse off by switching from m^i . If $w^i(s)R^i(s)b^i(s)$ then i is completely indifferent between all alternatives at s so m^i is not weakly dominated by any strategy. ■

Proof of Lemma 4:

If $s' \neq s$ then since $M_1 = \{(x, s) \mid x \in F(s)\}$, we trivially have $x \in F(s)$, so suppose $s \neq s'$ and $x \notin F(s')$. Then either (i) or (ii) of property Q (Definition 9') hold.

I. Suppose (i) holds. We claim that $\bar{m}^i = (x, s, s', -1, 2, s')$ weakly dominates m^i for agent $i_1(s, s')$. To see this, observe that the change from m^i to \bar{m}^i only affects the outcome when $(\bar{m}^{-i}, m^i) \in D_{4A}^i \cup D_{5A}^i \cup D_{8A}^i$. We analyze each case separately.

(1) If $(\bar{m}^{-i}, m^i) \in D_{4A}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i \cup D_{7A}^i$. The outcome changes from $y_1(s, s'')$ in D_{4A}^i to $b^i(s')$ in D_{7A}^i (since $k^j = 0 \forall j \neq i$) or from $y_1(s, s')$ to $y_2(s, s')$ in D_{6A}^i .

By definition, $b^i(s')R^i(s')y_1(s, s'')$ and $y_2R^i(s')y_1$.

(2) If $(\bar{m}^{-i}, m^i) \in D_{5A}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{7A}^i$, and the outcome changes from $w^i(s'')$ to $b^i(s')$ if $k^i > \max_{j \neq i} k^j$ and remains at $w^i(s'')$ otherwise. When $s'' = s'$ and $k^j = 1$

$\forall j \neq i$, i is strictly better off since by definition of $i_1(s, s')$, $b^i(s')R^i(s')w^i(s')$.

(3) $(\bar{m}^{-i}, m^i) \in D_{8A}^i \Rightarrow (\bar{m}^{-i}, \bar{m}^i) \in D_{8A}^i$. The outcome either remains at $w^i(s'')$ or changes to $b^i(s')$. In either case, i is no worse off.

Hence, if $i = i_1(s, s')$, i is no worse off anywhere and is strictly better off as described in (2), so \bar{m}^i weakly dominates m^i at s' .

II. Suppose (ii) holds. We claim that $\bar{m}^i = (x, s, s', -3, 2, s')$ weakly dominates m^i for agent $i_2(s, s')$. To see this, observe that the change from m^i to \bar{m}^i only affects the outcome when $(\bar{m}^{-i}, m^i) \in D_{4B}^i \cup D_{5B}^i$. We analyze each case separately.

(1) If $(\bar{m}^{-i}, m^i) \in D_{4B}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{6B}^i \cup D_{7B}^i$. When $(\bar{m}^{-i}, \bar{m}^i) \in D_{6B}^i$, then the outcome changes from $z_1(s, s')$ in D_{4B}^i to $z_2(s, s')$ in D_{6B}^i , and i is strictly better off since by definition, $z_2P^i(s')z_1$. If $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$, the outcome either changes from $z_1(s, s'')$ to $z^*(s', s, s'')$ or remains $z_1(s, s'')$. In either case, i is no worse off.

(2) If $(\bar{m}^{-i}, m^i) \in D_{5B}^i$, $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$, and the outcome either changes from $z_1(s, s'')$ to $z^*(s', s, s'')$ or remains at $z_1(s, s'')$. In either case, i is no worse off.

Hence, if $i = i_2(s, s')$, i is no worse off anywhere and is strictly better off as described in (4), so \bar{m}^i weakly dominates m^i at s' .

To conclude, if $x \notin F(s')$ then m^i is weakly dominated, so that if $m^j = (x, s, s, 0, 0, s) \forall j$ is a UNE at $s' \neq s$, then $x \in F(s')$. ■

Proof of Lemma 5:

As in Lemma 4, without loss of generality, we may assume that $s' \neq s$ and $x \notin F(s')$. Then either (i) or (ii) of Definition 9' hold.

If (i) of Definition 9' holds, consider agent $i = i_1(s, s')$:

(a) If i is playing $m^i = (x, s, s, 0, 0, s)$, then the same argument as in Lemma 4 can be used to show that m^i is weakly dominated.

(b) If i is playing $m^i = (x, s, s''', -1, k^i, s')$, we claim $\bar{m}^i = (x, s, s''', -1, k^i + 1, s')$

weakly dominates m^i . To see this, note that the change from m^i to \bar{m}^i only affects the outcome in D_{7A}^i and in D_{8A}^i . In D_{7A}^i , when $\bar{m}^j = (x, s, s', -2, k^i, s) \forall j \neq i$, the outcome changes from $w^i(s')$ to $b^i(s')$, and i is strictly better off. For any other \bar{m}^j , $j \neq i$, such that $(\bar{m}^{-i}, m^i) \in D_{7A}^i$, i is no worse off.

If $(\bar{m}^{-i}, m^i) \in D_{8A}^i$, the outcome can change from $w^i(s'')$ to $b^i(s')$ or remain the same. In either case, i is no worse off. Hence $g(\bar{m}^{-i}, \bar{m}^i)R^i(s')g(\bar{m}^{-i}, m^i) \forall \bar{m}^{-i}$ and $g(\bar{m}^{-i}, \bar{m}^i)P^i(s')g(\bar{m}^{-i}, m^i)$ when $\bar{m}^j = (x, s, s', -2, k^i, s) \forall j \neq i$, so that \bar{m}^i weakly dominates m^i .

(c) i is playing $m^i = (x, s, s''', -1, k^i, s')$ for $s^i \neq s'$. Again, changing from s^i to s' never hurts i , so \bar{m}^i given in (b) weakly dominates m^i .

(d) i is playing $m^i = (x, s, s''', -3, k^i, s')$. First, we claim that if

$z_2(s, \hat{s})P^i(s')z_1(s, \hat{s})$ for some \hat{s} such that $i = i_2(s, \hat{s})$, then

$\bar{m}^i = (x, s, s''', -3, k^i + 1, s')$ weakly dominates m^i . To see this, note that the change from m^i to \bar{m}^i only affects the outcome in D_{7B}^i . When $\bar{m}^j = (x, s, \hat{s}, -4, k^i, s) \forall j \neq i$,

the outcome switches from $z_1(s, \hat{s})$ to $z^*(s', s, \hat{s}) = z_2(s, \hat{s})$

since $z_2(s, \hat{s})P^i(s')z_1(s, \hat{s})$, and i is strictly better off. Further, i is no worse off for any \bar{m}^{-i} such that $(\bar{m}^{-i}, m^i) \in D_{7B}^i$. This establishes the claim.

Suppose, then, that $z_1(s, \hat{s})R^i(s')z_2(s, \hat{s})$ for all \hat{s} such that $i = i_2(s, \hat{s})$. In this case, $z^*(s', s, \hat{s}) = z_1(s, \hat{s})$ for all such \hat{s} , and we claim $\bar{m}^i = (x, s, s', -1, 2, s')$ weakly dominates m^i . To see this, observe that the outcome can

only change in $D_{4A}^i \cup \dots \cup D_{7B}^i$, and we handle each case separately:

(i) If $(\bar{m}^{-i}, m^i) \in D_{4A}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i \cup D_{7A}^i$. If $(\bar{m}^{-i}, \bar{m}^i) \in D_{6A}^i$, the outcome changes from $y_1(s, s')$ to $y_2(s, s')$, and by definition, $y_2R^i(s')y_1$, so i is no worse off. If $(\bar{m}^{-i}, \bar{m}^i) \in D_{7A}^i$, the outcome changes from $y_1(s, s'')$ to $b^i(s')$ since $k^j = 0 \forall j \neq i$, and i is no worse off.

(ii) If $(\bar{m}^{-i}, m^i) \in D_{5A}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{7A}^i$. When $\bar{m}^j = (x, s, s', -2, 1, s) \forall j \neq i$, the

outcome changes from $w^i(s')$ to $b^i(s')$ and i is strictly better off. For any other \bar{m}^{-i} such that $(\bar{m}^{-i}, \bar{m}^i) \in D_{7A}^i$, i is no worse off.

(iii) $(\bar{m}^{-i}, m^i) \notin D_{6A}^i \cup D_{7A}^i$

(iv) If $(\bar{m}^{-i}, m^i) \in D_{8A}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{8A}^i$, and the outcome either remains $w^i(s'')$ or changes to $b^i(s')$. In either case, i is no worse off.

(v) $(\bar{m}^{-i}, m^i) \notin D_{4B}^i \cup D_{5B}^i$

(vi) If $(\bar{m}^{-i}, m^i) \in D_{6B}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{4B}^i$, and the outcome changes from $z_2(s, s'')$ to $z_1(s, s'')$ and i is no worse off since by hypothesis $z_1(s, s'')R^i(s')z_2(s, s'')$.

(vii) If $(\bar{m}^{-i}, m^i) \in D_{7B}^i$ then $(\bar{m}^{-i}, \bar{m}^i) \in D_{4B}^i$ so the outcome remains at $z_1(s, s'')$.

To summarize, i is no worse off anywhere and strictly better off in (ii) above, so \bar{m}^i weakly dominates m^i .

(e) i is playing $m^i = (x, s, s''', -3, k^i, s')$, $s^i \neq s'$.

Again, changing from s^i to s' never hurts i at s' , and the argument in (d) applies.

To summarize, if (i) of Definition 9' holds, then m^i is weakly dominated.

Suppose then that (ii) of Definition 9' holds, and consider agent $i = i_2(s, s')$.

(f) i is playing $m^i = (x, s, s''', -3, k^i, s')$. We claim that $\bar{m}^i = (x, s, s''', -3, k^i + 2, s')$ weakly dominates m^i . To see this, note that the outcome

can only change in D_{7B}^i . Since $z_2(s, s')P^i(s')z_1(s, s')$, $z^*(s', s, s') = z_2(s, s')$, so when $\bar{m}^j = (x, s, s', -4, k^i + 1, s)$, the outcome changes from $z_1(s, s')$ to $z^*(s', s, s')$ and i is strictly better off. For any \bar{m}^{-i} such that $(\bar{m}^{-i}, \bar{m}^i) \in D_{7B}^i$, i is no worse off.

(g) i is playing $m^i = (x, s, s''', -3, k^i, s')$, $s^i \neq s'$. Again, switching from s^i to s' never hurts i , and the argument in (f) applies.

(h) i is playing $m^i = (x, s, s''', -1, k^i, s')$. Here, $\bar{m}^i = (x, s, s''', -1, k^i + 2, s')$ weakly dominates m^i . To see this, note that the outcome only changes in D_{8A}^i ,

i is strictly better off when $\bar{m}^j = (x, s, s', -2, k^i + 1, s) \forall j \neq i$, and i is no worse off anywhere else.

(i) Finally, if i is playing $m^i = (x, s, s'', -1, k^i, s^i)$, $s^i \neq s'$, switching s^i to s' never hurts i , and case (h) applies.

To conclude the proof of Lemma 5, if $x \notin F(s')$, then m^i is weakly dominated, so if $m \in D_2$ is a UNE at s' with $m_1^j = (x, s) \forall j$ then $x \in F(s')$. ■

REFERENCES

- Abreu, D. and A. Sen, "Subgame Perfect Implementation," Princeton University, 1986.
- Blume, L., "Lexicographic Refinements of Nash Equilibrium," University of Michigan, 1986.
- Brandenburger, A. and E. Dekel, "On an Axiomatic Approach to Refinements of Nash Equilibrium: A Characterization of Perfect Equilibrium," University of California - Berkeley, 1986.
- Crawford, V. "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations," *Econometrica* 47: 49-60, 1979.
- van Damme, E., *Refinements of the Nash Equilibrium Concept*, Springer-Verlag, Berlin, 1983.
- Dasgupta, P., Hammond, P. and E. Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies* 46: 185-216, 1979.
- Gibbard, A., "Manipulation of Voting Schemes: A General result," *Econometrica*, Vol. 41, pp. 587-601, 1973.
- Groves, T. and J. Ledyard, "Incentive Compatibility Ten Years Later," Northwestern University, 1985.
- Hurwicz, L., "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points," *Review of Economic Studies*, Vol. 46, pp. 217-225, 1979a.
- Hurwicz, L., "On Allocations Attainable Through Nash-Equilibria," in *Aggregation and Revelation of Preferences*, ed. J.-J. Laffont, pp. 397-419, North-Holland, Amsterdam, 1979b.
- Hurwicz, L., E. Maskin and A. Postlewaite, "Feasible Implementation of Social Choice Correspondences by Nash Equilibria," mimeo, 1984.
- Maskin, E., "Nash Equilibrium and Welfare Optimality," 1977, *Mathematics of Operations Research* (forthcoming).
- Maskin, E., "The Theory of Implementation in Nash Equilibrium: A Survey", in *Social Goals and Social Organization: Volume in Memory of Elisha Pazner*, Cambridge University Press, 1986.
- Matsuo, R., "The Design of Games for Implementation in Nash Equilibrium," University of Pennsylvania, 1985.
- McKelvey R., "Game Forms for Nash Implementation of General Social Choice Correspondences," Social Science Working Paper #579, Caltech, 1985.
- Moore, J. and R. Repullo, "Subgame Perfect Implementation," London School of Economics Working Paper, 1986.

- Moulin, H. "Dominance Solvable Voting Schemes," *Econometrica* 47: 1337-1352, 1979.
- Palfrey, T. and S. Srivastava, "Implementation with Incomplete Information in Exchange Economies," Carnegie-Mellon University, 1985.
- Palfrey, T. and S. Srivastava, "On Bayesian Implementable Allocations," 1986a, forthcoming *Review of Economic Studies*.
- Palfrey, T. and S. Srivastava, "Implementation in Exchange Economies Using Refinements of Nash Equilibrium," Carnegie-Mellon University, 1986b.
- Palfrey, T. and S. Srivastava, "Notes on Undominated Bayesian Implementation", Carnegie-Mellon University, 1986c.
- Postlewaite, A., "Implementation via Nash Equilibria in Economic Environments," in *Social Goals and Social Organization: A Volume in Memory of Elisha Pazner*, ed. L. Hurwicz, D. Schmeidler, and H. Sonnenschein, 1986.
- Postlewaite, A. and D. Schmeidler, "Implementation in Differential Information Economies," *Journal of Economic Theory*, Vol. 39, pp. 14-33, 1986.
- Saijo, T., *Nash Implementation Theory*, Ph.D. Dissertation, University of Minnesota 1985.
- Satterthwaite, M., "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, Vol. 10, pp. 187-217, 1975.
- Schmeidler, D., "Walrasian Analysis via Strategic Outcome Functions," *Econometrica*, Vol. 48, pp. 1585-1593, 1980.
- Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, Vol. 4, pp. 25-55, 1975.
- Walker, M., "A Simple Incentive Compatible Scheme for Attaining Lindahl Allocations," *Econometrica*, Vol. 49, pp. 65-73, 1981.
- Williams, S., "Sufficient Conditions for Nash Implementation," IMA preprint series, University of Minnesota, 1984.